Collection MICROLOG sur Micromédia
Tableau de résolution

Droits d'auteurs

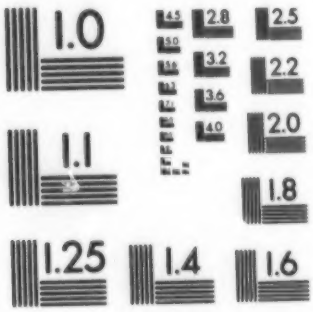
Les droits d'auteurs du conenu de cette microfiche sont détenus par l'éditeur du document original. Veuillez communiquer avec celui-ci pour obtenir des renseignements sur tout ce qui touche les droits d'auteurs. Les droits d'auteurs de cette édition sur microfiche sont détenus par Micromédia Limitée.

Qualité de reproduction sur microfiche

Micromédia Limitée maintient les normes de reproduction micrographique les plus élevées. Elle fait tout son possible pour obtenir la version sur papier à reproduire sur microfiche de la meilleure qualité possible. Cependant, certains styles d'imprimerie, de couleurs et de style de mise en page graphique (incluant des cartes, des grahiques et des tableaux) se reproduisent mal. En raison de la valeur du document original, Micromédia l'a reproduit tout en reconnaissant qu'il ne sera peut-être pas aussi lisable que le document original.

Tableau de résolution micrographique
National Bureau of Standards
Standard Reference Material 1010a
(ANSI and ISO Test Chart No. 2)

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT



MS109

24x

NATIONAL MICROGRAPHICS ASSOCIATION

micromedia
a division of IHS Canada

20 Victoria Street, Toronto, ON M5C 2N8
Tel.: (416) 362-5211
Toll free: 1-800-387-2689
Fax: (416) 362-6161

RECEIVED DEC 2 4 2001

Ocean Sciences Data Rescue – 1

Anthony W. Isenor, Inessa Yashayaeva and
Paula Willcott

Ocean Sciences Division
Maritimes Region
Fisheries and Oceans Canada

Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, Nova Scotia
Canada B2Y 4A2

2001

Canadian Technical Report of Hydrography and Ocean Sciences 217

Fisheries and Oceans Canada | Pêches et Océans Canada

Canada

102-01676/

Canadian Technical Report of Hydrography and Ocean Sciences 217

2001

Ocean Sciences Data Rescue - I

by

Anthony W. Isenor, Inessa Yashayaeva[1] and Paula Willcott[2]

Ocean Sciences Division
Martimes Region
Fisheries and Oceans Canada
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth Nova Scotia,
Canada
B2Y 4A2

[1]BDR Research Ltd.
Box 652, Stn Central,
Halifax, NS, B3J 2T3

[2]Evans Computing Ltd.
6424 Norwood Street
Halifax, Nova Scotia
Canada, B3H 2L3

© Public Works and Government Services 2001
Cat. No. Fs 97-18/217 E ISSN: 0771-6764

Correct citation for this publication:

Isenor, Anthony W., Inessa Yashayaeva and Paula Willcott. 2001. Ocean Sciences Data Rescue - I, Can. Tech. Rep. Hydrog. Ocean Sci. 217, iv + 29p.

Table of Contents

ABSTRACT IV

RÉSUMÉ IV

1.0 INTRODUCTION 1

2.0 PRE-RESCUE CHECKS 2

2.1 Data Source 2
2.2 Cruise Check Procedure 2

3.0 PROCEDURES 3

3.1 Bridge Log and Cruise Reports 3
3.2 CTD Data from Wet and Dry Deck Sheets 4
3.3 Thermometer Temperature Computations 5
3.4 Water Sample Salinity 5
3.5 Water Sample Oxygen and Nutrient Data 7
3.6 Pressure to Depth Conversion 8
3.7 Quality Control 8

4.0 RESCUED CRUISE DATASETS 10

4.1 Rescued Datasets 10
4.2 Typical Rescue Problems 10
4.3 Port to BIOCHEM 13
4.4 Paper Record Archival 13

5.0 CONCLUSIONS 13

6.0 ACKNOWLEDGEMENTS 14

7.0 REFERENCES 14

APPENDIX A: NUMBER OF DIGITS DELIVERED WITH THE DATA TYPES 15

APPENDIX B: DIGITIZING AND PROCESSING STEPS 16

APPENDIX C: BIOCHEM PORT FORMAT 17

APPENDIX D: IMPORT ERROR REPORTS 21

Abstract

Isenor, Anthony W., Inessa Yashayaeva and Paula Willcott. 2001. Ocean Sciences Data Rescue - I, Can. Tech. Rep. Hydrog. Ocean Sci. 217, iv + 29p.

Water sample data consisting of thermometer temperature, salinity, oxygen and nutrients and supporting metadata have been digitised from historic records and placed in a local biological and chemical oceanographic database. The digitised data were quality controlled by plotting and visually comparing values within a parameter type and dataset. A transfer format was developed between the digitising environment and the target database. In total, 27 datasets representing over 1200 profiles were digitised and transferred to the database. These data did not previously exist in a form accessible by researchers.

Résumé

Isenor, Anthony W., Inessa Yashayaeva and Paula Willcott. 2001. Ocean Sciences Data Rescue - I, Can. Tech. Rep. Hydrog. Ocean Sci. 217, iv + 29p.

On a numérisé des données historiques d'échantillonnage de l'eau - portant sur la température thermométrique, sur la salinité, sur l'oxygène et sur les nutriments - ainsi que les métadonnées justificatives, et on les placées dans une base locale de données biologiques et d'océanographie chimique. La qualité des données numérisées a été contrôlée par tracé graphique et comparaison visuelle des valeurs correspondant à un type de paramètre et à un ensemble de données. Un format de transfert de l'environnement de numérisation à la base de données de destination a été élaboré. En tout, 27 ensembles de données, représentant plus de 1 200 profils, ont été numérisés et transférés à la base de données. Ces données n'existaient pas jusqu'ici dans une forme accessible aux chercheurs.

1.0 INTRODUCTION

Advances in computer technology combined with electronic databases are providing a new method of storing, processing and disseminating data to users. In oceanography, these databases are housing historic data of significant importance to the research and environmental science communities. However, gaps in the datasets are being detected, by comparing past collection exercises to the contents of the data archives.

These data gaps are troubling to the clients of the databases. The collected data is arguably our most valuable resource and is being recognised as such by many business organisations. In science, we are also recognising the importance of these historic data sets.

The historic datasets also represent a considerable investment. At present, one day of ship time for a deep ocean research vessel is approximately $20,000 Cdn. In the 1970s, a comparable cost would have been $4300 (simply using inflation index from the consumer price index). Another important point is with regard to the replacement of the historic datasets. One should realise that historic datasets cannot be replaced with any amount of funds. We cannot re-measure values from previous times.

The above points indicate obvious advantages to including all historic datasets within the databases designed to house these data. In terms of data management, there are also advantages to keeping these data in one system, such as maintenance and scale-ability.

Internationally there is considerable effort being expended toward rescue of historic datasets. The Intergovernmental Oceanographic Commission (IOC) Global Ocean Data and Archaeology Rescue (GODAR) project has been underway since 1993. Recent reports suggest that this project alone has been responsible for the rescue of 190,000 conductivity-temperature-depth (CTD) profiles, 1.5 million bottle stations and 21,000 profiles of biological data (ICES, 2000).

In terms of Canadian data rescue, the Marine Environmental Data Service (MEDS) and the National Data Committee have recognised the need for data rescue. A National Data Management Policy recognises data rescue as a key component in the preservation of data for departmental programs.

Locally, the Ocean Sciences Division (OSD) at the Bedford Institute of Oceanography (BIO) has been very active in seeking out and restoring historic datasets. A recent report (Isenor and Woldegeorgis, 2000) identified all science cruises from BIO. Tracking data collected on the cruises from Ocean Sciences Division, a comparison of collected data and data existing in the archive databases identified numerous missing datasets. The project described here is a step toward actually rescuing some of these identified datasets.

This report outlines this OSD data rescue effort. The primary source of information is reviewed. Problems with a rescue effort are identified. The details of the data processing and transfer to the existing data archive are discussed to assist others involved in similar projects.

2.0 PRE-RESCUE CHECKS

In data rescue projects, the greatest fear is spending time rescuing a dataset that you later identify as existing in quality digital form. In an attempt to avoid this unfortunate situation, we implemented a series of procedures and checks to determine the existence of a cruise dataset within the archive databases. These procedures are outlined in this section.

A frequent problem encountered during the rescue process is scattered data files. Cruise data can frequently be distributed over many locations, each containing some small piece of the larger dataset. For example, water sample log sheets may be in the national archives, while file folders containing the data are in various filing cabinets around the Institute. Typically, the data are with those scientists interested in that particular data type, with orphan data types often stored in the OSD Datashop.

2.1 Data Source

Initially, the primary source of historic records for this rescue project was local paper archives stored within Ocean Circulation Section. As the project evolved, paper records from the BIO library and digital records from Marine Environmental Sciences Divisions (MESD) were also utilised.

As a means of organising the rescue effort, the records were first catalogued. All relevant file folders, printouts, binders, etc. were identified and inventoried. All information was associated with a cruise number, thus allowing easy searches for specific cruise related questions.

The records contained in the Ocean Circulation storage room and the Ocean Sciences Datashop were not itemised in detail. The cruise numbers displayed on the filing cabinets were listed and cabinets searched based on the cruise number of interest.

2.2 Cruise Check Procedure

At the time of this rescue project, a major database development project involving the creating and filling of the biological and chemical oceanographic database (BIOCHEM) was underway at BIO. BIOCHEM represented the combination of about 15 biological and chemical databases within OSD and MESD. Previous to BIOCHEM, the MESD cruise and nutrient databases were the main archive destinations for water sample data.

By searching the BIOCHEM (or equivalent) database, we identified existing digital datasets. The database searches concentrated on both the database cruise number field and for any string containing the cruise number in the comment field. This was required due to the possible identification of an archived dataset by more than one cruise number.

Multiple cruise numbers for a single cruise dataset may occur due to multiple pathways for a dataset. Datasets collected by BIO scientists are shared internationally. Through data sharing

agreements, these datasets may return to BIO databases. In some cases, the returned datasets may contain altered cruise numbers. Within BIOCHEM, if these data are identified as originating from a BIO cruise, the BIO cruise number is included in the comment section for the dataset. On occasion, the cruise number was insufficient for proper identification. In this case, a spatial-temporal search for stations was considered to determine if a dataset was present in the archive database.

The OSD data archive was also identified as a potential source of data. For some cruise datasets, the water sample data were discovered in the OSD archive that did not exist in BIOCHEM.

The BIO library was also searched for historic data reports from sources such as the Canadian Oceanographic Data Centre (predecessor of MEDS). All reports were inventoried for water sample oxygen and nutrient data along with accompanying metadata.

Finally, we also had a digital list of cruises that did not exist in the BIOCHEM database, but were thought to have collected nutrient data. This list, compiled by Strain and Bussard (Strain, pers. comm.), was based on a search of the BIO library cruise report archives. This list provided some guidance as to data we should be looking for.

3.0 PROCEDURES

A proper data rescue project requires well-defined rescue procedures. For this project, emphasis was placed on rescuing water sample salinity, oxygen and nutrient data. These data also require supporting data and metadata. Supporting data included the CTD pressure, temperature and salinity. These values will allow the conversion of units if necessary. Supporting metadata includes station position, time, and sounding.

A common problem during the rescue was the number of digits to deliver with the various data types. Calculations (e.g. pressure to depth) often result in an excess of digits after the decimal point. To address this, delivered numeric data were examined and the number of digits after the decimal point was specified. These numbers are given in Appendix A.

The following outlines the procedures followed during this rescue project.

3.1 Bridge Log and Cruise Reports

The science bridge log and cruise report from a particular cruise represent a valuable source of information for a data rescue project. However, this information is valuable from the standpoint of descriptive information, not necessarily a source of metadata or data. Often, bridge logs from these cruises datasets were difficult to interpret, containing multiple position entries for a single station, without clear identification relating the individual position to a particular stage of the station.

Cruise reports are also valuable information sources, but not particularly valuable data sources. This is expected, given the general nature of these reports. Occasionally, listings of station positions within the reports did represent valuable metadata.

3.2 CTD Data from Wet and Dry Deck Sheets

On typical BIO cruises involving the CTD system, water samples are collected in rosette bottles that surround the frame used to house the electronic equipment. A typical CTD cast involves a computer that acts as a logging device for the electronic data. Associated with the cast are paper records that note important position and time information. These paper records, known as "deck sheets" exist in both the computer room (dry deck sheets) and the winch room (wet deck sheets).

The deck sheets provide much of the metadata for the CTD cast. The time, date, sounding and position are recorded on the sheets. As well, the deck sheets provide the link between the BIO sample ID numbers and the rosette bottles. The sample ID numbers are uniquely numbered adhesive labels attached to the rosette bottle. These numbers provide a unique database key for the water sample data.

The deck sheets also provide CTD information from the time of rosette bottle trip. The CTD pressure, temperature, and conductivity readings are all recorded on the dry deck sheet when a bottle is remotely closed from the computer room control area.

For present day operations, dry deck sheets record CTD data in engineering units. However, the historic deck sheets often contain non-standard units. For example, the CTD pressure data from the dry deck sheets were typically in the form of percent full scale. Lazier (pers. comm.) indicated that the CTD pressure full-scale value was 6000 dbar. Some deck sheets actually provided pressure values in the comment section of the deck sheet. The 6000-dbar full scale value was verified using these comment notes and knowledge of the ocean depth for particular station locations.

CTD pressure percent full-scale values were digitised directly from the dry deck sheets. The pressure was then computed based on the 6000 dbar full scale value for the pressure.

The CTD temperature values did not require any further calculations. CTD temperature in ITPS-68 scale was digitised directly from the dry deck sheets. However, as with the CTD salinities digitised from the deck sheets, these values will not have had any calibration applied. Thus, all CTD data delivered as part of this rescue are uncalibrated.

CTD conductivities were also digitised from the deck sheet. Standard formulas (Fofonoff and Millard, 1983) were used to compute the in-situ CTD salinity given the CTD pressure, temperature and conductivity at the time of bottle trip.

The dry deck sheets also provided the station position and date/time information. For some datasets, the time and position information were obtained from the CTD profile data existing in

the OSD CTD archive. Station number was then used to link the time and position from the CTD profiles to the deck sheet sample ID numbers.

The wet deck sheets were used to obtain the sounding data. As well, wet deck sheets provided the reversing thermometer readings. Comments from the deck sheets regarding bottle trips and other such information were digitised in the Excel spreadsheet but were not transferred to the BIOCHEM database.

3.3 Thermometer Temperature Computations

Numerous cruise datasets contained data from reversing thermometers. These data were digitised by entering the sample ID number, thermometer serial number, and main and auxiliary temperature readings into an Excel spreadsheet. The data were then exported to a Paradox compatible file. The WOCE Toolbox (Isenor and Jackson, 1997) was used to compute actual thermometer temperature and pressure (when applicable).

The temperature computation used the thermometer calibration data from the closest, previous calibration. In this case, the calibration data would be the most recent available at the time of cruise sailing. It is difficult to assess whether or not this is the most applicable calibration information to use. Later calibrations may have been considered more appropriate and used without sufficient documentation noting this decision.

For some datasets, average temperatures were available from paper records as well as the unprocessed thermometer readings from the wet deck sheets. In one of these cases, a comparison was made between the paper record averages with those averages computed based on wet deck sheet information. This comparison is shown in Figure 1.

A total of 242 of the 258 comparisons are shown. The missing 16 comparisons are larger than ±0.03. A total of 40% of the values are clustered at 0 with an additional 23% at ±0.01. We consider this fair agreement, considering the complexities of the comparison. For example the historic records may be based on a calibration applied via look up tables for a particular thermometer and applied interpolation. As well, the exact calibration information used is uncertain.

3.4 Water Sample Salinity

In some cases, conductivity values from water samples were digitised from the conductivity log sheets. For many of the cruise datasets from the 1980's, the standards used were based on chlorinity.

For digitising, we included the sequence number on the log sheet, the water sample ID number, the Autosal temperature, and the water sample conductivity. The sequence number was used to ensure the entered values were in the order as defined on the conductivity log sheet. This order

M121

IMPORTANT NOTE CONCERNING THE FOLLOWING PAGES

THE PAGES WHICH FOLLOW HAVE BEEN FILMED TWICE IN ORDER TO OBTAIN THE BEST REPRODUCTIVE QUALITY

USERS SHOULD CONSULT ALL THE PAGES REPRODUCED ON THE FICHE IN ORDER TO OBTAIN A COMPLETE READING OF THE TEXT.

REMARQUE IMPORTANTE CONCERNANT LES PAGES QUI SUIVENT

LES PAGES SUIVANTES ONT ÉTÉ REPRODUITES EN DOUBLE AFIN D'AMÉLIORER LA QUALITÉ DE REPRODUCTION

LES UTILISATEURS DOIVENT CONSULTER TOUTES LES PAGES REPRODUITES SUR LA FICHE AFIN D'OBTENIR LA LECTURE DU TEXTE INTÉGRAL

micromedia
a division of IHS Canada

20 Victoria Street
Toronto, Ontario M5C 2N8
Tel.: (416) 362-5211
Toll Free: 1-800-387-2689
Fax: (416) 362-6161
Email: info@micromedia.on.ca

is important for the calculations during those instances when the Autosal instrument drifted over the completion of the samples.

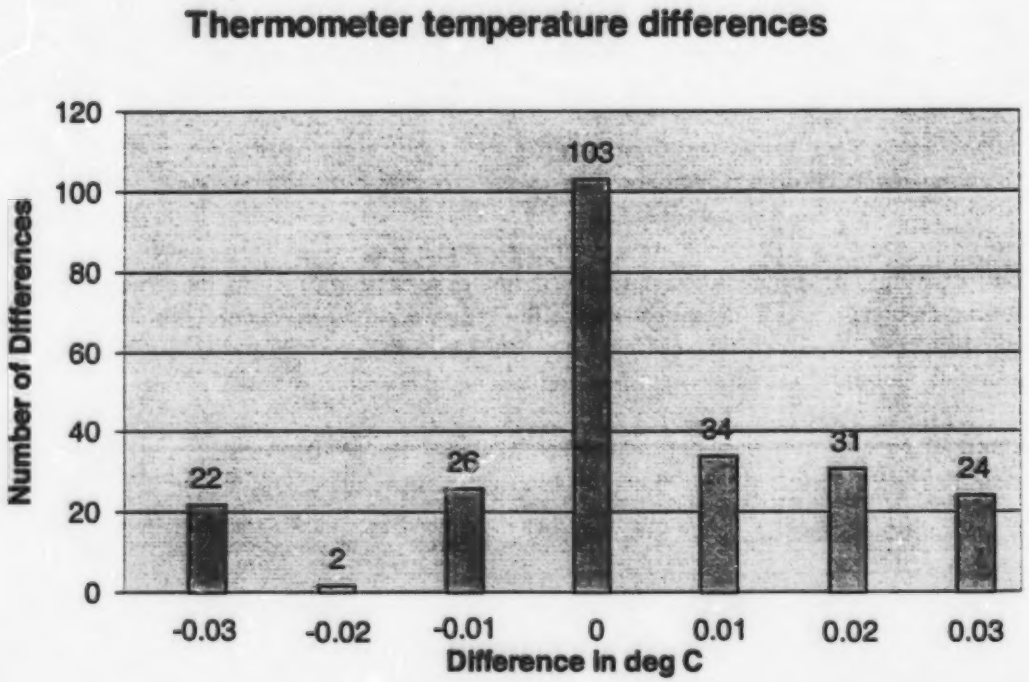


Figure 1. Histogram showing differences between average thermometer temperatures from paper records and averages computed during this rescue effort.

The digitising resulted in Excel spreadsheets containing the above parameters. The chlorinity of the standard water was converted to salinity using the formula (Pickard and Emery, 1985):

$$S_s = \text{Chlorinity} * 1.80655$$

The standard water salinity S_s was then converted to conductivity using standard formulas (Fofonoff and Millard, 1983) with values of P = 0 dbar and T = 15°C. The pressure-temperature values used in this conversion represent standard-temperature-pressure (STP), not the lab values. This is because the standard water has a determined chlorinity at STP.

For the water samples, the conductivity (C) was converted to twice the conductivity ratio (C_R).

$$C_R = 2 * (C / 42.914)$$

(42.914 is the conductivity at STP (P=0, T=15, S=35) expressed in Siemens/metre)

This produced a conductivity ratio similar to that produced by present day Autosal salinometers.

The data were then exported to Paradox compatible tables. The WOCE Toolbox (Isenor and Jackson, 1997) was then used to compute the water sample salinity. The WOCE Toolbox follows Fofonoff and Millard (1983) to compute salinity from conductivity ratio.

As a check on the computations, a comparison based on two cruise datasets (82038 and 85004) was conducted. The water sample salinities were computed based on the above procedure and compared with water sample salinities as found in the paper records for that cruise. The entire salinity dataset provided 266 comparison points. A histogram of the difference between the salinities from paper records and the computed salinities is shown in Figure 2.

The figure shows six differences of ±0.002 and 105 differences in the range of ±0.001. Without knowing the exact method of the historic computation, we consider this an acceptable range for the differences.

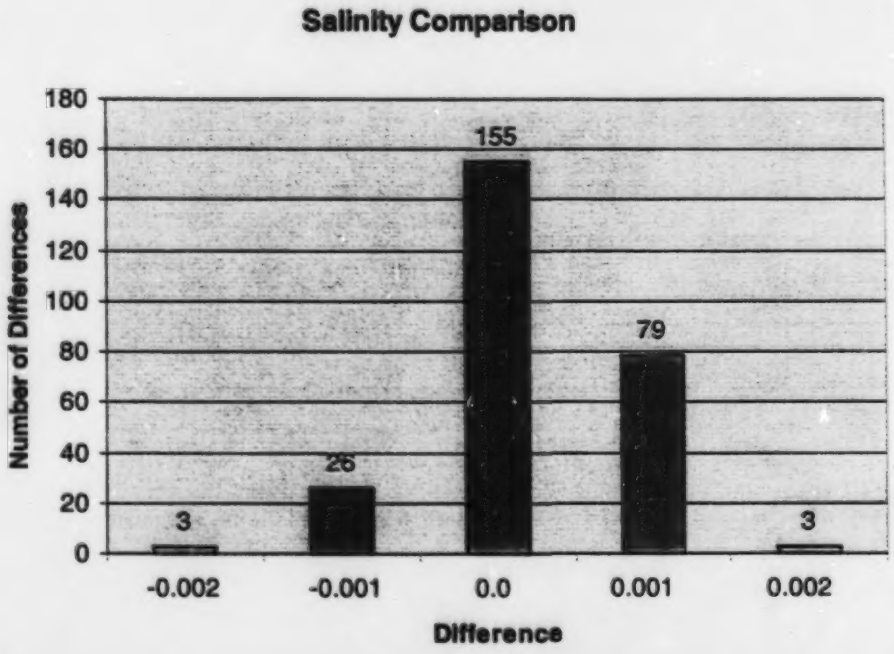


Figure 2. Histogram showing differences between the water sample salinities from paper records compared to those salinities computed based on conductivity log sheets. These data are a combination of differences from cruises 82038 and 85004.

3.5 Water Sample Oxygen and Nutrient Data

All paper log sheets of water sample oxygen and nutrient data were digitised. Typically, unit information was unavailable. We assume the units for oxygen are ml/l and for nutrients micro

is important for the calculations during those instances when the Autosal instrument drifted over the completion of the samples.

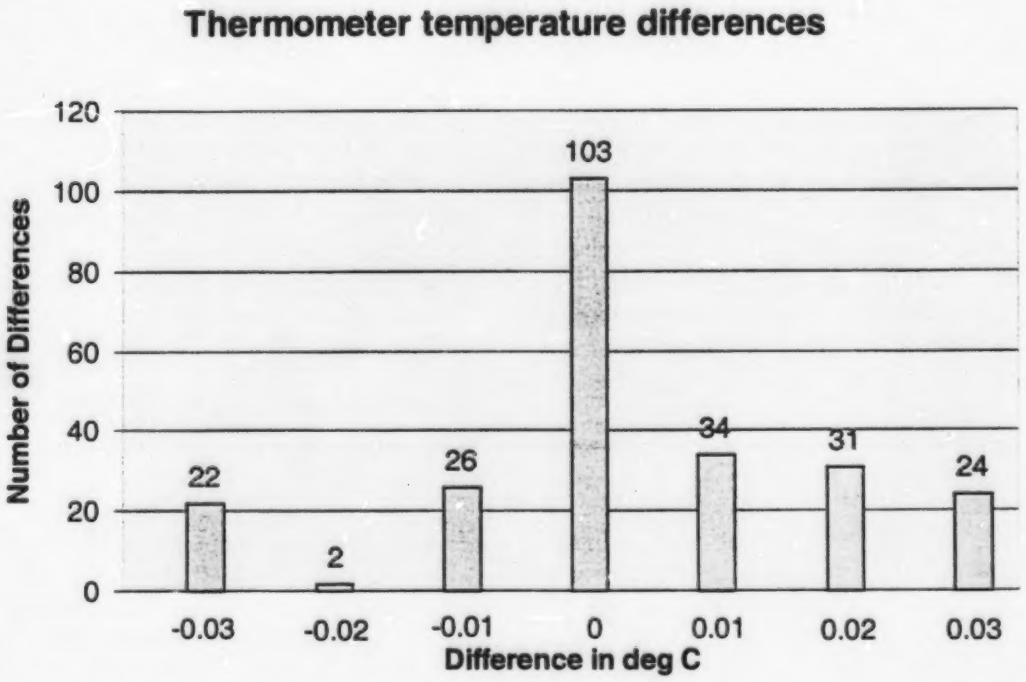


Figure 1. Histogram showing differences between average thermometer temperatures from paper records and averages computed during this rescue effort.

The digitising resulted in Excel spreadsheets containing the above parameters. The chlorinity of the standard water was converted to salinity using the formula (Pickard and Emery, 1985):

$$S_s = \text{Chlorinity} * 1.80655$$

The standard water salinity S_s was then converted to conductivity using standard formulas (Fofonoff and Millard, 1983) with values of P = 0 dbar and T = 15°C. The pressure-temperature values used in this conversion represent standard-temperature-pressure (STP), not the lab values. This is because the standard water has a determined chlorinity at STP.

For the water samples, the conductivity (C) was converted to twice the conductivity ratio (C_R).

$$C_R = 2 * (C / 42.914)$$

(42.914 is the conductivity at STP (P=0, T=15, S=35) expressed in Siemens/metre)

This produced a conductivity ratio similar to that produced by present day Autosal salinometers.

The data were then exported to Paradox compatible tables. The WOCE Toolbox (Isenor and Jackson, 1997) was then used to compute the water sample salinity. The WOCE Toolbox follows Fofonoff and Millard (1983) to compute salinity from conductivity ratio.

As a check on the computations, a comparison based on two cruise datasets (82038 and 85004) was conducted. The water sample salinities were computed based on the above procedure and compared with water sample salinities as found in the paper records for that cruise. The entire salinity dataset provided 266 comparison points. A histogram of the difference between the salinities from paper records and the computed salinities is shown in Figure 2.

The figure shows six differences of ±0.002 and 105 differences in the range of ±0.001. Without knowing the exact method of the historic computation, we consider this an acceptable range for the differences.

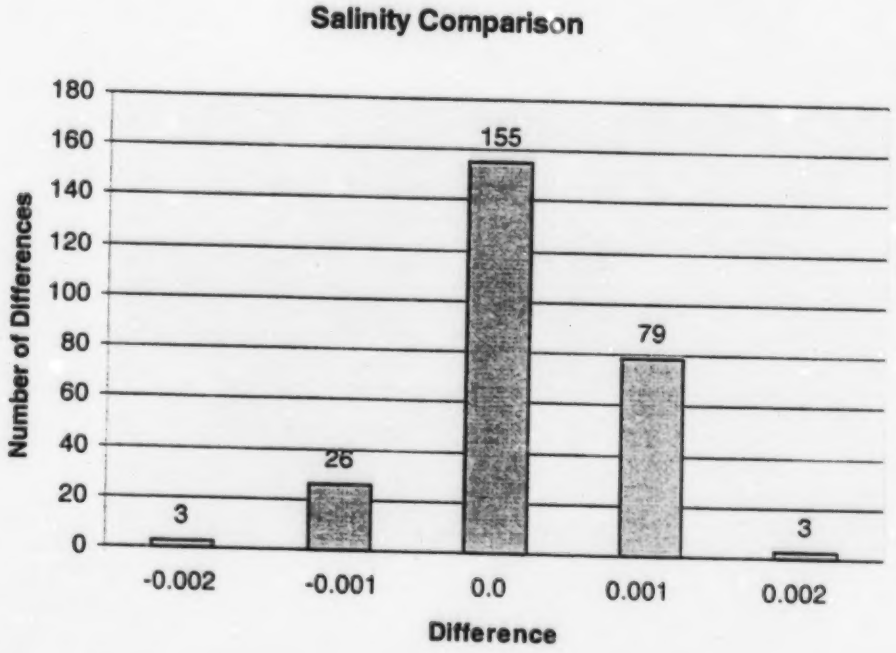


Figure 2. Histogram showing differences between the water sample salinities from paper records compared to those salinities computed based on conductivity log sheets. These data are a combination of differences from cruises 82038 and 85004.

3.5 Water Sample Oxygen and Nutrient Data

All paper log sheets of water sample oxygen and nutrient data were digitised. Typically, unit information was unavailable. We assume the units for oxygen are ml/l and for nutrients micro

moles/l. These are valid assumptions because the historic apparatus used to determine the values reported the data in these units.

For some datasets, oxygen values were calculated from titer volumes. Historically, titer volumes were recorded on titration sheets. For several datasets, these titration sheets were digitised and oxygen values calculated. Calculations were performed according to Jones et al. (1992). Fortunately, the titration records also contained other necessary information such as flask number, blank information and normalities of added solutions.

Some nutrient data included in this rescue were already available in digital form from MESD. These data had not previously been incorporated into the larger dataset. For this rescue we incorporated these electronic data into the cruise dataset.

3.6 Pressure to Depth Conversion

As noted above, the dry deck sheet information contains the CTD pressure which is used to locate the water samples in the vertical. However, the BIOCHEM data model requires the depth of the sample rather than the pressure.

To convert pressure to depth we applied the formula from Saunders and Fofonoff (1976). Input for this conversion is pressure and latitude. For this conversion, we used the station latitude. The output depth was then rounded to the nearest metre. This accounts for the accuracy of the pressure transducer which is reported to be 0.1 % full scale (online documentation for pressure transducers used on Guildeline CTDs at http://www.viatran.com/103/x18specs.htm), or about six dbars full scale (full scale is 6000 dbars).

3.7 Quality Control

Overall, we consider the flagging conducted during this rescue to be very poor. The visual display of the water sample data, typically a parameter plotted against pressure but sometimes against another parameter, can only result in the most basic of quality control. Nevertheless, this process does help to identify extreme data points within the series.

This type of quality control is prone to errors in the flagging. This is especially true with parameter-parameter plots. In this case, one parameter may be the cause of the outlier while both parameters become flagged as a result of the interpretation.

Problems were enhanced by the initial plotting procedures used during this project. The initial plotting, which was performed in Excel, required human manipulation to convert from plot space to data and flag values within the data file. As well, this environment did not allow true zooming within the plot region. The plot region could be magnified, however, this does not change the plot scale. This results in an inability to separate data points in a cluster. If the cluster of points should have a bad or questionable flag assigned, then the Excel plotting technique and data point identification will result in points being missed within the cluster.

These problems initiated the use of in-house quality control software called QCPLOT. This software is a Matlab based plotting and flagging tool specifically developed for discrete sample data. QCPLOT allows one to retrieve data and flags from a database table, and visually set the data flags. The flags are then replaced in the database table. Full zoom capabilities are present in QCPLOT, thus allowing full manipulation over the scale and data display. This tool greatly reduced quality control processing time.

As an example of the applied quality control, consider Figure 3. Here, a plot of water sample oxygen data from cruise 90022 is plotted against pressure. The figure shows points flagged according to MEDS definitions, and using the following symbols:

(•) value appears to be correct - MEDS flag 1
(+) value is doubtful - MEDS flag 3
(*) value appears to be wrong - MEDS flag 4

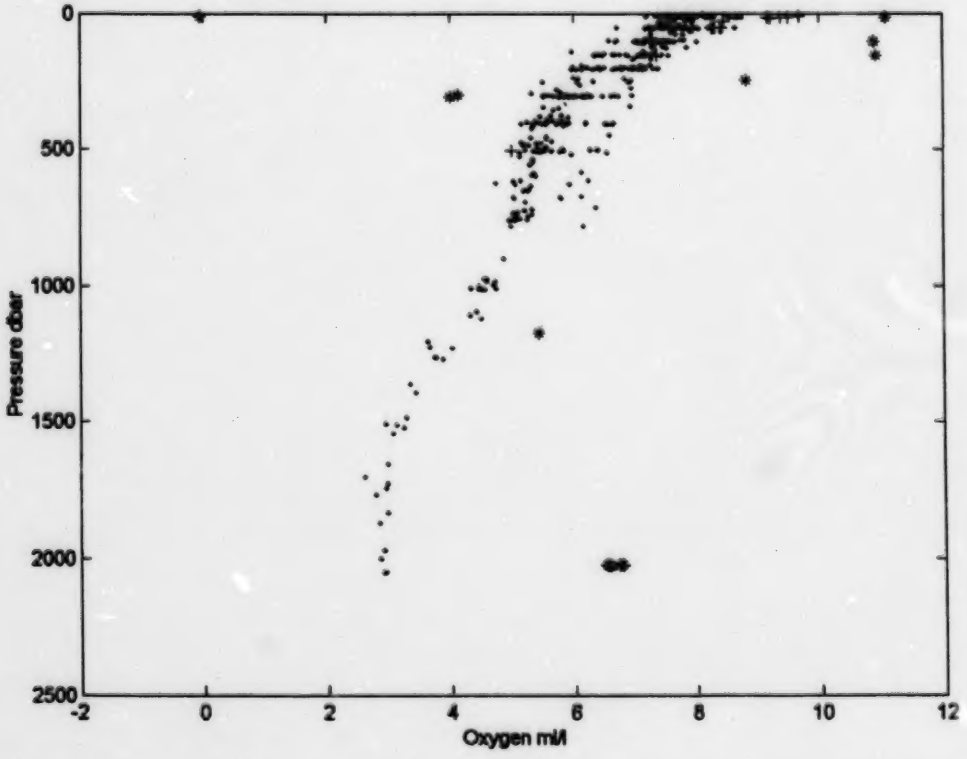


Figure 3. Water sample oxygen plotted against pressure for cruise 90022. The type of symbol used in the plot indicates the flags. Flags are correct (•), doubtful (+) and wrong (*).

4.0 Rescued Cruise Datasets

After following the procedure to identify a missing cruise dataset, we began searching the two main locations for available data. It should be noted that many of these cruises were multidisciplinary, however, the paper records existing within Ocean Circulation typically only pertained to physical and chemical data. A summary of the digitising and processing steps is presented in Appendix B.

4.1 Rescued Datasets

Table 1 provides a summary of the datasets rescued during this project. The table is sorted on ascending cruise number. The other columns indicate the presence of water sample salinity, oxygen, nutrients and thermometer temperature in the dataset.

Of the 27 rescued datasets, only 41% were identified as missing by Isenor and Woldegeorgis (2000). If this ratio is typical of the under reporting, then we could expect the 86 datasets identified by Isenor and Woldegeorgis (2000) to be an underestimate, with the actual number being closer to 200.

Table 2 provides a summary of rescued data points by parameter category. The standard parameters of salinity and temperature dominate. However, over 6000 water sample oxygen and nutrient values were recovered and are valuable additions to the archives.

4.2 Typical Rescue Problems

There are many problems associated with the rescue of datasets. Most of these problems are organisational in nature.

Probably the most serious general problem relates to the source information being scattered among numerous locations. It was typical for cruise information to be located in the OSD datashop, Divisional filing cabinets, MESD and national archives. This scattering of information makes identification of a complete dataset time consuming and difficult. This scattering also points to the need for inventories of the various locations, to allow one to track the parts of the dataset that appear in these locations.

A second general problem relates to the inconsistent state of information. Essentially, the information may be present in any form, ranging from an individual's logbook to digital. Often these inconsistent formats result in inconsistent relationships between data types. Established relationship structures, such as the BIO sample ID number system, is not necessarily present across all data forms.

The most serious specific dataset problem was encountered with the 86029 cruise dataset. Arguably, the most important component of the BIO water sample system is the sample ID number. A unique sequential number is assigned to each rosette bottle, and all subsamples from

that rosette bottle are tracked based on this ID number. For this particular dataset, five different ID numbering sets were used. This makes the joining of different parameter types difficult.

Overall, these problems point to either the lack of established procedures or the non-compliance with established procedures.

Cruise Number	Salinity	Oxygen	Nutrients	Thermometer Temperature
67002	y	y	y	y
76020	y	y	y	
79022	y	y	y	y
80029	y	y		y
81035	y			y
82038	y	y		y
83021	y			y
83024	y			y
83029	y	y		y
83032	y			y
83036	y	y		y
84007	y			y
84023	y			y
84026	y			y
84036	y	y		y
84038	y			y
85004	y	y		y
85018	y	y	y	y
86021/1	y	y	y	y
86021/2	y			y
86029	y		y	y
87002	y			y
87031	y			y
88025	y	y		y
89016/2	y	y	y	y
89037	y	y	y	
90022	y	y	y	y

Table 1. The rescued cruise datasets. Columns indicate the cruise number and the presence of water sample salinity, oxygen, nutrients and thermometer temperature in the dataset ("y" indicates "Yes" it was present, blank cells indicate that no data of that type could be found).

Cruise Number	No. of Events with Data	No. of Input Records	No. of Salinity values	No. of Oxygen values	No. of Nutrient values	No. of CFC 12 values	No. of CFC 11 values	No. of Therm. Readings
67002	119	12257	2379	2070	2354*			2376
76020	46	1672	330	329	343*			
79022	55	4142	210	403	981			91
80029	16	488	95	100				73
81035	15 (2)	516	68					197
82038	28	596	80	30				167
83021	32 (1)	530	61					177
83024	30	285	56					88
83029	35	3671	779	779				524
83032	7	207	27					55
83036	56	1805	219	150				473
84007	18	465	68					114
84023	44	279	91					49
84026	36	873	144					374
84036	29	3494	624	627				431
84038	23 (1)	484	64					180
85004	38	4426	841	271				643
85018	36 (2)	1497	200	156	177*			186
86021/1	23	2503	247	149	147*	107	108	32
86021/2	70	939	180					183
86029	56	1658	176		443*			97
87002	89	683	84					255
87031	176	2099	356					537
88025	46	1748	328	329				188
89016/2	16	2761	50	157	568*			38
89037	38	2889	353	358	346*			
90022	114	5554	198	556	824			558
TOTAL	**1297**	**58521**	**8308**	**6464**	**6183**	**107**	**108**	**8086**

Table 2. Total number of individual parameter values rescued within this project. The number of input records (includes discrete CTD data at the time of bottle trip) and the number of profile events containing data are also given. The number of unusable events is indicated with parentheses [e.g. (2) unusable events for cruise 81035]. All replicates are included in the numbers, with the exception of temperature, where the numbers given do not include duplicate readings. For nutrients, an asterisk (*) is used to indicate that no replicates are present.

4.3 Port to BIOCHEM

The port to BIOCHEM had two main problems.

The first problem was the unpredictable date format specifications within Excel. Changes to cell date formats were sometimes successful, but often were not successful. The reasons for this are unclear. For this reason, the focus of the data entry and processing shifted from Excel to Access during the course of the project.

Second, the BIOCHEM import procedures were evolving as this rescue took place. This resulted in changes to our delivered data structure through the course of the rescue. This complicated the transfer and increased the number of iterations required to find an acceptable format. However, once the format and procedure were defined, the transfer was reasonably straightforward.

This testing of port format specifications represents a useful outcome from this rescue project. For this reason, the BIOCHEM port format is detailed in Appendix C. The Appendix outlines the Access table format required for the data transfer in this project.

The results of the data port to BIOCHEM are outlined in Appendix D. Initial error reports produced during the import into BIOCHEM, are given for all datasets. Import errors were subsequently corrected and the data were re-imported.

4.4 Paper Record Archival

There was considerable effort made in assembling the various paper records for the data and metadata. To take advantage of this effort, we centralised and packaged the assembled records into standard units suitable for delivery to the national archive. With the permission of the chief scientist these records were passed to the OSD Datashop for delivery to the national archives.

5.0 CONCLUSIONS

Considerable resources have been spent to collect oceanographic data from areas of interest to Canadian researchers. Occasionally, the collected data do not get placed in the final databases designed to store and provide these data to clients. In this situation, it is important to recognise the important contribution these data may make to the historic records and attempt to identify and recover these data.

This report described in detail the rescue of cruise water sample data and supporting metadata. In total, 27 cruise datasets were rescued representing about 29000 combined water sample values of salinity, thermometer temperature, oxygen and nutrients. The procedures used to rescue the data were described and some problems with such a process were noted.

The tasks associated with data rescue are not glamorous. However, the funds needed to support such rescue efforts are quite small compared to present-day collection costs. Management must recognise the cost-effective nature of such work and support it both in spirit and in funding.

6.0 ACKNOWLEDGEMENTS

A special thanks to Peter Strain and Robert Benjamin for their help in defining a port format and for their efforts and patience to read our data files. Jeffery Jackson developed the QCPLOT software, which greatly reduced the time spent quality controlling the water sample data. This work was supported by funding from the Marine Environmental Data Service, the Department of Fisheries and Oceans Strategic Science Fund Ocean Climate Program and the Panel for Energy, Research and Development.

7.0 REFERENCES

Fofonoff, N.P. and R.C. Millard Jr. 1983. Algorithms for computation of fundamental properties of seawater, UNESCO technical paper in marine science 44.

Isenor, Anthony W. and Fascil Woldegeorgis. 2000. Ocean Sciences Data Rescue Inventory, Can. Tech. Rep. Hydrog. Ocean Sci. 205, iv + 30p.

Isenor, A. W. and J. Jackson. 1997. WOCE Toolbox User's Guide, unpublished manuscript.

ICES, 2000. Report of the Working Group on Marine Data Management, Hamburg Germany.

Jones, E.P., F. Zemlyak and P. Stewart. 1992. Operating Manual for the Bedford Institute of Oceanography Automated Dissolved Oxygen Titration System. Can. Tech. Rep. of Hydrography and Ocean Sci. 138: iv + 51 p.

Pickard, George L. and William J. Emery. 1985, Descriptive Physical Oceanography An Introduction, Pergamon Press, 249 pp.

Saunders, P.M. and N.P. Fofonoff. 1976. Conversion of pressure to depth in the ocean, Deep-Sea Res., 23, 109-111.

Appendix A: Number of Digits Delivered with the Data Types

CTD salinity - 4 (computed) or 3 (decksheets)	e.g. 34.1234
CTD temperature - determined by data on the deck sheets	
CTD pressure - 0	e.g. 3241 (pressure sensor repeatability is 0.1% full scale. See online documentation for pressure transducers used on Guildeline CTDs at http://www.viatran.com/103/x18specs.htm).
Water sample salinity - 3	e.g. 34.123
Thermometer temperature - 2	e.g. 2.12
Thermometer pressure - 0	e.g. 3241
Water sample oxygen - 3	e.g. 3.123
Depth - 0	e.g. 3241 (pressure sensor is only good to 0.1% full scale)
Latitude - 5	e.g. 42.12345 (this gives > 0.001 minute resolution)
Longitude - 5	e.g. 42.12345

Appendix B: Digitizing and Processing Steps

Station information from wet and dry deck sheets was digitised into Excel spreadsheets or Access tables. CTD discrete data was obtained from dry deck sheets, thermometer readings from wet deck sheets, and water sample salinity, oxygen and nutrients from logbooks or other such paper records. On some occasions, data were available in electronic form.

If only conductivity log sheets were available, water sample salinities were computed using the WOCE Toolbox.

Thermometer temperatures were also computed using the WOCE Toolbox. The calibration information available at the time of cruise sailing was used to compute these temperatures (and pressures if applicable).

CTD discrete salinity data were computed from CTD conductivity, pressure and temperature using Excel functions developed within OSD. The pressure, in percent full-scale, was first converted to dbars.

Initial Excel Procedure:

The data types (water sample and CTD) were then placed in Access and queries run to join the data using sample ID number.

Results of the query went back to Excel for plotting and quality control. Quality flags were assigned to the data based on visual checks.

Excel modules were then developed to convert the "horizontal" dataset (each sheet particular to a data type) into "vertical" dataset (BIOCHEM parameter number and value).

The data were moved to Microsoft Access and queries run to combine the metadata with the data and produce a BIOCHEM port format.

Data were then moved back to Excel to combine cruise-station-sample identifiers and convert dates to a particular format.

Revised Access Procedure:

The data types (water sample and CTD) were imported into Access.

QCPLOT was used to visually examine the data and set quality flags as necessary.

Access queries were developed to combine standard input data and metadata tables and output a BIOCHEM port format (given in Appendix C).

Appendix C: BIOCHEM Port Format

The following describes the port format between the data rescue effort and the BIOCHEM data model. Excel spreadsheet files were initially produced in the following format and provided to the BIOCHEM data managers. However, this procedure was later revised such that the following format was provided in Access tables. The data contained in the provided files were imported into BIOCHEM where validation took place.

Note that the following are presented in transposed form. Thus the BIOCHEM fields are actually columns in the delivered sheets or tables.

Station Format:

BIOCHEM Field name	BIOCHEM Variable Type	Description	Access Data Type
DIS_SAMPLE_KEY_VALUE	VARCHAR2	This is a sample counter. For modern day data, this could be just the sample ID number. A safe way of dealing with this is to assign the number like cruise_station_sampleID. This must match the corresponding field in the Data file. The _station_ ensures uniqueness when there are more than one station without ID numbers.	text
MISSION_DESCRIPTOR	VARCHAR2	Cruise number. Needs to be the same as DATA file.	text
EVENT_COLLECTOR_EVENT_ID	VARCHAR2	For CTD or bottle casts, this is the station number. Needs to be the same as DATA file.	text
EVENT_COLLECTOR_STN_NAME	VARCHAR2	Must be same as in DAT file. This is now a required field. This may be set equal to the EVENT_COLLECTOR_EVENT_ID	text
MISSION_NAME	VARCHAR2		text
MISSION_LEADER	VARCHAR2	Name of Chief Scientist	text
MISSION_SDATE	DATE	Start Date of cruise. Can obtain this from the BCD.	date/time
MISSION_EDATE	DATE	End Date of cruise. Can obtain this from the BCD.	date/time
MISSION_INSTITUTE	VARCHAR2	This should be: BIO	text
MISSION_PLATFORM	VARCHAR2	Ship name. Obtain from BCD.	text
MISSION_PROTOCOL	VARCHAR2		text
MISSION_GEOGRAPHIC_REGION	VARCHAR2	Location. Obtain from BCD.	text
MISSION_COLLECTOR_COMMENT1	VARCHAR2		text
MISSION_COLLECTOR_COMMENT2	VARCHAR2		text
MISSION_DATA_MANAGER_COMMENT	VARCHAR2		text
EVENT_SDATE	DATE	Station start date.	date/time

EVENT_EDATE	DATE	Station end date. If we don't have this value, it can be the same as the start date.	date/time
EVENT_STIME	NUMBER	Station start time. Must be in GMT. Format is hhmm.	number long integer
EVENT_ETIME	NUMBER	Station end time. If we don't have this value, it can be the same as the EVENT_STIME. Must be in GMT. Format is hhmm.	number long integer
EVENT_MIN_LAT	NUMBER	Station latitude value.	number double
EVENT_MAX_LAT	NUMBER	Station latitude value.	number double
EVENT_MIN_LON	NUMBER	Station longitude value.	number double
EVENT_MAX_LON	NUMBER	Station longitude value.	number double
EVENT_UTC_OFFSET	NUMBER		number double
EVENT_COLLECTOR_COMMENT1	VARCHAR2	This field should contain: "Assumed speed of sound for sounding is 800 fathoms/sec"	text
EVENT_COLLECTOR_COMMENT2	VARCHAR2		text
EVENT_DATA_MANAGER_COMMENT	VARCHAR2		text
DIS_HEADR_GEAR_SEQ	NUMBER	Sampler type used to collect the sample. See below.	number double
DIS_HEADR_SDATE	DATE	If possible, this should be the date of bottle trip. If not, the station date should be used. Needs to be the same as DATA file.	date/time
DIS_HEADR_EDATE	DATE	For bottle trips, this should be the same as EVENT_EDATE.	date/time
DIS_HEADR_STIME	NUMBER	If possible, this should be the time of bottle trip. If not, the EVENT_STIME should be used. Needs to be the same as DATA file.	number long integer
DIS_HEADR_ETIME	NUMBER	For bottle trips, this should be the same as EVENT_STIME.	number long integer
DIS_HEADR_TIME_QC_CODE	VARCHAR2		text
DIS_HEADR_SLAT	NUMBER	Station latitude. Needs to be the same as DATA file.	number double
DIS_HEADR_ELAT	NUMBER		number double
DIS_HEADR_SLON	NUMBER	Station longitude. Needs to be the same as DATA file.	number double
DIS_HEADR_ELON	NUMBER		number double
DIS_HEADR_POSITION_QC_CODE	VARCHAR2		text
DIS_HEADR_START_DEPTH	NUMBER	If possible, we should convert the pressure of the bottle trip to depth.	number double
DIS_HEADR_END_DEPTH	NUMBER	Same as start depth. This must match the DIS_HEADER_END_DEPTH column in the DAT table.	number double

DIS_HEADR_SOUNDING	NUMBER	Station sounding in metres.	number double
DIS_HEADR_COLLECTOR_DEPLMT_ID	VARCHAR2		text
DIS_HEADR_COLLECTOR_SAMPLE_ID	VARCHAR2	Sample ID number.	text
DIS_HEADR_COLLECTOR	VARCHAR2		text
DIS_HEADR_COLLECTOR_COMMENT1	VARCHAR2		text
DIS_HEADR_DATA_MANAGER_COMMENT	VARCHAR2	This comment should read: "These data were rescued as part of the Ocean Circulation Data Rescue Project, A. Isenor, 2000."	text
DIS_HEADR_RESPONSIBLE_GROUP	VARCHAR2	This should be "MESD and OSD at BIO"	text
DIS_HEADR_SHARED_DATA	VARCHAR2		text
CREATED_BY	VARCHAR2	Should read", "Data Rescue 1 - AI"	text
CREATED_DATE	DATE	Todays date.	date/time
DATA_CENTER_CODE	NUMBER	This should be: 20	number double
PROCESS_FLAG	VARCHAR2	This should be "NR"	text
BATCH_SEQ	NUMBER		number double

* The CREATED_BY filed read "Data Rescue 2 - A. Isenor" for Part II of the rescue effort.

Data Format:

BIOCHEM Field Name	BIOCHEM Variable Type	Description	Access Data Type
MISSION_DESCRIPTOR	VARCHAR2	Cruise number. Must be same as in STN file.	text
EVENT_COLLECTOR_EVENT_ID	VARCHAR2	This is station number for a CTD or bottle cast. Must be same as in STN file.	text
EVENT_COLLECTOR_STN_NAME	VARCHAR2	Must be same as in STN file. This is now a required field. This may be set equal to the EVENT_COLLECTOR_EVENT_ID	text
DIS_HEADER_START_DEPTH	NUMBER	If possible, we should convert the pressure of the bottle trip to depth. Must be same as in STN file.	number double
DIS_HEADER_END_DEPTH	NUMBER	Same as start depth. Must be same as in STN file.	number double
DIS_HEADER_SLAT	NUMBER	Station latitude. Must be same as in STN file.	number double
DIS_HEADER_SLON	NUMBER	Station longitude. Must be same as in STN file.	number double
DIS_HEADER_SDATE	DATE	If possible, this should be the date of bottle trip. If not, the station date should be used. Must be same as in STN file.	date/time

DIS_HEADER_STIME	NUMBER	If possible, this should be the time of bottle trip. If not, the station time should be used. Must be same as in STN file.	number long integer
DIS_DETAIL_DATA_TYPE_SEQ	NUMBER	Parameter number. See below for parameter numbers that we will typically use.	number long integer
DATA_TYPE_METHOD	VARCHAR2	This should be blank, no quotes.	text
DIS_DETAIL_DATA_VALUE	NUMBER	The parameter value.	number double
DIS_DETAIL_DATA_QC_CODE	VARCHAR2	MEDS flags.	text
DIS_DETAIL_DETECTION_LIMIT	NUMBER		number double
DIS_DETAIL_DETAIL_COLLECTOR	VARCHAR2		text
DIS_DETAIL_COLLECTOR_SAMP_ID	VARCHAR2	Sample ID Number	text
CREATED_BY	VARCHAR2	Should read", "Data Rescue 1 - AI"	text
CREATED_DATE	DATE	Todays date.	date/time
DATA_CENTER_CODE	NUMBER	This should be the number 20.	number double
PROCESS_FLAG	VARCHAR2	This should be "NR"	text
BATCH_SEQ	NUMBER		number double
DIS_SAMPLE_KEY_VALUE	VARCHAR2	This is a sample counter. For modern day data, this could be just the sample ID number. A safe way of dealing with this is to assign the number like cruise_sampleID. This should match the corresponding field in the Station file.	text

Parameter Numbers used in BIOCHEM and this rescue:

Parameter	Number
CTD Pressure	90000001
CTD Temperature	90000004
CTD Salinity	90000104
CTD Oxygen	90000009
Autosal salinity	90000105
oxygen - water sample	90000007
silicate - water sample	90000018
NO2+NO3 - water sample	90000020
phosphate - water sample	90000019
Thermometer temperature	90000005
Thermometer Pressure	90000124
Freon 11	90000072
Freon 113	90000083
Freon 12	90000073
Total Phosphorus	90000080

Appendix D: Import Error Reports

The following error reports are broken into two parts, representing the Excel verses Access procedure (see Appendix B). Both parts are then broken into actual shipments of datasets. A shipment represents a package of data assembled and delivered to the BIOCHEM data manager. The ***bold italic*** indicates the reply to the error report, and any suggested correction.

Legend:

filename

- Station records transferred (number of records in the station table)
- Data records transferred (number of records in the data table)
- Errors (any reported errors)
- Mission (number of missions. This will always be one)
- Events (number of stations, with and without data present)
- Headers (number of usable records in the station table)
- Detail (number of unique IDs with data)
- Replicates (number of replicate ID records with data)
- Errors (any reported errors)

Part I - First Shipment:

80029_RB.xls

- Created date not equal in station and data tables
- Station Created_date = 8/11/00 10:16:00 AM
- Data Created_date = 8/11/00 10:33:00 AM
- Changed Created_date in both tables to equal 8/11/00 10:16:00 AM

```
UPDATE Data SET Data.CREATED_DATE = #8/11/00 10:16:00 AM#
WHERE (((Data.CREATED_DATE)=#8/11/00 10:33:00 AM#));
```

- 106 Station records transferred
- 488 Data records transferred
- Mission 1
- Events 16
- Headers 106
- Detail 421
- Replicates 119
- Errors

  Sample_id 802519 missing Depth value - ***(we have no information on these IDs)***

82038_RB.xls

- 98 Station records transferred
- 596 Data records transferred
- Mission 1
- Events 28
- Headers 94
- Detail 468
- Replicates 223
- Errors Sample_id 4683 and 4684 missing depth values. - ***(we have no information on these IDs)***

83036_RB.xls

- 241 Station records transferred
- 1805 Data records transferred
- Mission 1
- Events 56
- Headers 240
- Detail 1416
- Replicates 720
- No Errors

87031_RB.xls

- 349 Station records transferred
- 2099 Data records transferred
- Mission 1
- Events 176
- Headers 349
- Detail 1722
- Replicates 674
- No Errors

89037_RB.xls

- 380 Station records transferred
- 2889 Data records transferred
- Mission 1
- Events 38
- Headers 380
- Detail 2889
- Replicates 0
- No Errors

Total Number of events: 314

Part I - Second Shipment:

83021_RB.xls

- 73 Station records transferred
- 530 Data records transferred
- Mission 1
- Events 33
- Headers 73
- Detail 391 Errors
  - Sample_id 1040 missing value QC code – ***value QC should be 1***
- Replicates 235
- Errors
  - Sample_id 1062, 1063, 1064, 1065 (Position QC = 4) - ***Position QC should be 1 for this station***
  - Sample_id 1028, 1029 missing depth - ***(we have no information on these IDs)***

85018_PB.xls

- 186 Station records transferred
- 1497 Data records transferred
- Mission 1
- Events 38
- Headers 152
- Detail 1321
- Replicates 292
- Errors
  - Sample_id 4201 4202 4203 4204 4205 (Position QC = 4) – ***Longitude value should be –55.88983 and Position QC=1 for this station***
  - Sample_id 4053 4160 4221 4222 4223 4224 4225 4226 4227 4228 4229 4230 4231 4232 4233 4234 (Missing depth) - ***(we have no information on these IDs)***

86021_1RB.xls

- 256 Station records transferred
- 2503 Data records transferred
- Mission 1
- Events 23
- Headers 244
- Detail 1957
- Replicates 1076
- Errors
  - Sample_id 20421 - 20430, 20641 - 20657 (Position QC = 4)

Sample_id 20421 – 20430 - ***Longitude value should be –48.84057 and Position QC =1 for this station***

Sample_id 20641 – 20657 - ***Position QC should be 3***

20401 20577 20586 20587 20588 20601 20603 20630 20631 20632 20633 20634 (Missing depth) - ***(we have no information on these IDs)***

86021_2RB.xls

- 182 Station records transferred
- 939 Data records transferred
- Mission 1
- Events 70
- Headers 182
- Detail 805
- Replicates 213
- Errors

Sample_id 20824 Pressure = -1

- value QC (DIS_DETAIL_DATA_QC_CODE) should be 4 and the Depth value (DIS_HEADR_START_DEPTH & DIS_HEADR_END_DEPTH) should be –1 as well (in both Stations and Data worksheets). This is the value given on the deck sheets.

88025_RB.xls

- 296 Station records transferred
- 1748 Data records transferred
- Mission 1
- Events 46
- Headers 296
- Detail 1507
- Replicates 366
- Errors

Sample_id 49194 (Missing depth) - ***(we have no information on this ID)***

Total Number of events: 210

Part I - Third Shipment:

84007_RB.xls

- 70 Station records transferred
- 465 Data records transferred
- Mission 1
- Events 18
- Headers 70
- Detail 352
- Replicates 175
- No Errors

90022_RB.xls

- 773 Station records transferred
- 5554 Data records transferred
- Mission 1
- Events 114
- Headers 773
- Detail 4147
- Replicates 2760
- No Errors

84023_RB.xls

- 45 Station records transferred
- 279 Data records transferred
- Mission 1
- Events 44
- Headers 45
- Detail 218
- Replicates 116
- No Errors

84026_RB.xls

- 144 Station records transferred
- 873 Data records transferred
- Mission 1
- Events 36
- Headers 138
- Detail 678
- Replicates 310
- Errors Sample Id 11401, 11402, 11419, 11420, 11507, 11508 (Missing depth)) - ***(we have no information on these IDs)***

84038_RB.xls

- 65 Station records transferred
- 484 Data records transferred
- Mission 1
- Events 24
- Headers 63
- Detail 363
- Replicates 219
- Errors Sample Id 8411615, 8411616 (Missing Depth)) - ***(we have no information on these IDs)***

Total Number of events: 236

Part I - Fourth Shipment:

81035_RB.xls

- 68 Station records transferred
- 516 Data records transferred
- Mission 1
- Events 17 Error Event 9 - End time before start time - ***corrected and resubmitted. This also uncovered a processing error in BIOCHEM, where the date was not taken into account when comparing start and end times.***
- Headers 60
- Detail 370
- Replicates 250
- Error Missing Depths Sample_id 8705 to 8708 and 8745 to 8748

83024_RB.xls

- 60 Station records transferred
- 285 Data records transferred
- Mission 1
- Events 30
- Headers 57
- Detail 248
- Replicates 57
- Error Missing Depths Sample_id 1503 1512 1536

83029_RB.xls

- 728 Station records transferred
- 3671 Data records transferred
- Mission 1
- Events 35
- Headers 727
- Detail 3234
- Replicates 741
- Error Missing Depths Sample_id 16678

83032_RB.xls

- 28 Station records transferred
- 207 Data records transferred
- Mission 1
- Events 7
- Headers 28
- Detail 143
- Replicates 91
- No Errors

84036_RB.xls

- 608 Station records transferred
- 3494 Data records transferred
- Mission 1
- Events 29
- Headers 608
- Detail 3078
- Replicates 724
- Error Invalid start ne Sample_id 13727 - ***corrected and resubmitted.***

Part I - Fifth Shipment:

85004_RB.xls

- 797 Station records transferred
- 4426 Data records transferred
- Mission 1
- Events 38
- Headers 797
- Detail 3840
- Replicates 1017
- No Errors

87002_RB.xls

- 89 Station records transferred
- 683 Data records transferred
- Mission 1
- Events 89
- Headers 89
- Detail 520
- Replicates 303
- No Errors

Part II - First Shipment:

79022.mdb

- 495 Station records transferred
- 4142 Data records transferred
- Mission 1
- Events 55
- Headers 495
- Detail 2571
- Replicates 3142
- No Errors

86029.mdb

- 497 Station records transferred
- 1658 Data records transferred
- Mission 1
- Events 56
- Headers 490
- Detail 1584
- Replicates 186
- No Errors

89016.mdb

- 292 Station records transferred
- 2761 Data records transferred
- Mission 1
- Events 16
- Headers 292
- Detail 1917
- Replicates 1790
- No Errors

Part II - Second Shipment:

67002.mdb

- 2388 Station records transferred
- 12257 Data records transferred
- Mission 1
- Events 119
- Headers 2388
- Detail 11776
- Replicates 962
- No Errors

76020.mdb

- 347 Station records transferred
- 1672 Data records transferred
- Mission 1
- Events 46
- Headers 347
- Detail 1671
- Replicates 0
- No Errors